PATENT OFFICE
JAPANESE COVERNMENT

This is to certify that the annexed is a true copy of the following application as filed with this Office.

Date of Application: July 15, 1999
Application Number : P11-201988
Applicant(s) : KDD CORPORATION

December 10, 1999

Commissioner,
Patent Office Takahiko KONDOU

H11-3085579

日 本 国 特 許 庁
PATENT OFFICE
JAPANESE GOVERNMENT

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日
Date of Application: １９９９年 ７月１５日

出 願 番 号
Application Number: 平成１１年特許願第２０１９８８号

出 願 人
Applicant (s): ケイディディ株式会社

１９９９年１２月１０日

特 許 庁 長 官
Commissioner,
Patent Office 近 藤 隆 彦

出証番号 出証特平１１－３０８５５７９

【書類名】　特許願

【整理番号】　KDD-32

【提出日】　平成11年 7月15日

【あて先】　特許庁長官殿

【国際特許分類】　G06F 17/00

【発明の名称】　情報自動フィルタリング方法および装置

【請求項の数】　4

【発明者】

【住所又は居所】　埼玉県上福岡市大原２－１－１５　株式会社ケイディディ研究所内

【氏名】　井ノ上　直己

【発明者】

【住所又は居所】　埼玉県上福岡市大原２－１－１５　株式会社ケイディディ研究所内

【氏名】　帆足　啓一郎

【発明者】

【住所又は居所】　埼玉県上福岡市大原２－１－１５　株式会社ケイディディ研究所内

【氏名】　橋本　和夫

【特許出願人】

【識別番号】　000001214

【氏名又は名称】　ケイディディ株式会社

【代理人】

【識別番号】　100083806

【弁理士】

【氏名又は名称】　三好　秀和

【電話番号】　03-3504-3075

【選任した代理人】

【識別番号】　100100712

【弁理士】

【氏名又は名称】　岩▲崎▼　幸邦

【選任した代理人】

【識別番号】　100095500

【弁理士】

【氏名又は名称】　伊藤　正和

【選任した代理人】

【識別番号】　100101247

【弁理士】

【氏名又は名称】　高橋　俊一

【手数料の表示】

【予納台帳番号】　001982

【納付金額】　21,000円

【提出物件の目録】

【物件名】　明細書　1

【物件名】　図面　1

【物件名】　要約書　1

【包括委任状番号】　9506777

【プルーフの要否】　要

【書類名】　明細書

【発明の名称】　情報自動フィルタリング方法および装置

【特許請求の範囲】

【請求項１】　インターネットを介して提供される情報のうち不適切情報を識別し、この識別した不適切情報の提供を阻止する情報自動フィルタリング方法であって、

提供の阻止を必要とする不適切な情報および提供の阻止を必要としない適切な情報を学習データとした自動学習により前記情報に含まれる単語に対して情報の提供を阻止する必要があるか否かを判定するために使用される単語の重みを求め、

この求めた単語の重みを各単語に対応して重み付き単語リストとして記憶管理しておき、

インターネットを介して提供される情報を入力し、この情報に含まれる単語を抽出し、

この抽出した単語の各々に対する重みを前記重み付き単語リストから読み出し、

この読み出した各単語の重みの総和を算出し、この算出した総和に基づき前記情報の提供を阻止すべきか否かを判定すること

を特徴とする情報自動フィルタリング方法。

【請求項２】　前記単語の重みを求める処理は、前記不適切な情報と適切な情報に対してベクトル空間上で弁別できる線形識別関数に基づく自動学習により単語の重みを求めることを特徴とする請求項１記載の情報自動フィルタリング方法。

【請求項３】　インターネットを介して提供される情報のうち不適切情報を識別し、この識別した不適切情報の提供を阻止する情報自動フィルタリング装置であって、

提供の阻止を必要とする不適切な情報および提供の阻止を必要としない適切な情報を学習データとした自動学習により前記情報に含まれる単語に対して情報の提供を阻止する必要があるか否かを判定するために使用される単語の重みを求め

る単語重み学習手段と、

この求めた単語の重みを各単語に対応して重み付き単語リストとして記憶管理する重み付き単語リスト格納手段と、

インターネットを介して提供される情報を入力する入力手段と、

この入力された情報に含まれる単語を抽出する単語抽出手段と、

この抽出した単語の各々に対する重みを前記重み付き単語リストから読み出し、この読み出した各単語の重みの総和を算出し、この算出した総和に基づき前記情報の提供を阻止すべきか否かを判定する判定手段と

を有することを特徴とする情報自動フィルタリング装置。

【請求項４】　前記単語重み学習手段は、前記不適切な情報と適切な情報に対してベクトル空間上で弁別できる線形識別関数に基づく自動学習により単語の重みを求める手段を有することを特徴とする請求項３記載の情報自動フィルタリング装置。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】

本発明は、インターネットを介して提供されるメールを含む各種情報に対して、情報に出現する単語を抽出し、この単語に基づいて前記情報が不適切であるか否かを判定し、不適切な情報の提供を阻止する情報自動フィルタリング方法および装置に関する。

【０００２】

【従来の技術】

インターネットの急速な広がりに伴い、限られた専門家の道具でしかなかったコンピュータはごく一般の家庭や学校などにも導入され始めている。このため、これまでコンピュータに触れることすらなかった多くの一般人でも気軽にインターネットにアクセスすることが可能になった。こうした背景の中、近年深刻な問題となっているのがインターネット上に氾濫するポルノ画像などの有害情報に対する子供のアクセスである。この問題に対処するため、アメリカでは政府機関がインターネット上の情報を検閲することを可能にした「通信品位法」という法律

が提案されたが、裁判の結果、表現の自由を保証する憲法に違反すると判決され、立法することができなかった。

【０００３】

そこで最近注目されているのが「情報フィルタリング」という技術である。情報フィルタリングとは、ユーザがインターネット上の情報にアクセスする際にその情報の有害性をチェックし、有害と判定された場合は何らかの手段によりその情報へのアクセスをブロックするという技術である。

【０００４】

現在市販されている有害情報フィルタリングソフトで取り入れられている手法は大きく以下の４つに分類される。

【０００５】

（１）自己判定によるフィルタリング

（２）第三者の判定によるフィルタリング

（３）自動フィルタリング

（４）単語に対するスコア（得点）を利用する方式

ここではこの４つの手法について簡単に解説する。まず、自己判定によるフィルタリング手法ではWWW情報の提供者が自らのコンテンツの有害性について判定を行い、その結果をＨＴＭＬファイル内に記述する。フィルタリングソフトはこの記述された結果を参照し、有害と判断された場合にアクセスをブロックする。この手法によるフィルタリングを図４に示す。

【０００６】

図４に示す自己判定に基づくフィルタリングでは、米国マサチューセッツ工科大学のWorld Wide Web Consortium が作成したＰＩＣＳ（Platform for Internet Content Selection ）と呼ばれるインターネットコンテンツの評価を記述するための基準を使用している。ＰＩＣＳを使用することにより、コンテンツ提供者は簡単に自分の提供している情報を描写し、開示することができる。

【０００７】

多くの場合、コンテンツ提供者がこのような評価結果を公開する際には、ＰＩＣＳによる評価結果を出力する評価機関のサービスを利用する。このような評価

機関の代表として、Recreational Software Advisory Council（ＲＳＡＣ）やSafeSurfといった団体があげられ、それぞれ独自に設定した基準による評価結果を提供している。コンテンツ提供者はこれらの機関からの評価結果をＨＴＭＬファイルのヘッダに記述する。図５にこの評価結果の記述例を示す。

【０００８】

この自己判定はコンテンツ提供者の自主性に任せられるというのが現状である。そのため、多くのコンテンツ提供者がこの判定を受けようという意志を持たない限りは本手法による有効な有害情報フィルタリングは不可能であるといえる。

【０００９】

次に第三者による判定に基づくフィルタリングについて説明する。有害情報フィルタリングソフトを作成している業者の中には、WWW上のホームページの有害性を独自に判定し、その結果をフィルタリングソフトの判断基準とする手法を取り入れている。一般的には、この評価の結果として有害なホームページのＵＲＬ一覧が構築されている。このＵＲＬのリストはフィルタリングソフトと共にユーザに分配され、フィルタリングソフトの判断基準となる。多くの場合、フィルタリングソフトはこの有害ＵＲＬ一覧を定期的にダウンロードする仕組みになっている。第三者による判定に基づく有害情報フィルタリングの仕組みを図６に示す。

【００１０】

このような仕組みを持つソフトウェアの代表的なものとしてCyberPatrolがあげられる。CyberPatrolは「暴力」「性行為」など１３個のジャンルに対し、それぞれ有害ＵＲＬ一覧を持っており、これらのシステムに従って有害情報フィルタリングを行う。

【００１１】

この手法で使用される有害ＵＲＬ一覧はそれぞれソフトウェア業者でホームページをアクセスし、判定を行うことによって作成・拡張されているため、新しく設立されたホームページや従来のＵＲＬから別のＵＲＬに移動したホームページには対処することは不可能である。従って、こうした評価対象外のページに対するフィルタリングには対処できないのが現状である。

【００１２】

次に、自動フィルタリングについて説明する。有害情報フィルタリングソフトの中にはアクセスされたホームページの中身をチェックし、有害性の判断を行うものもある。

【００１３】

具体的には、有害な情報、すなわち不適切な情報内に含まれるであろう単語を予め登録しておき、この登録した単語が情報内に出現するか否かをチェックし、前記登録した単語が含まれていた場合に情報の提供を阻止する方式である。例えば、ポルノ情報の提供を阻止する場合、情報内に”ｓｅｘ”や“ｘｘｘ”といった文字列が含まれていた場合、その情報の提供を阻止する。この手法の応用として、登録した単語が情報内に含まれている割合が所定の閾値を上回った場合に情報の提供を阻止する方式もある。

【００１４】

次に、単語に対するスコア（得点）を利用する方式について説明する。この方式は、不適切な情報内に含まれるであろう単語およびこの単語に対するスコアを予め登録しておき、この登録した単語が情報内に出現するか否かをチェックし、登録した単語が含まれていた場合に単語のスコアを合計し、この合計が所定の閾値を上回った場合に該情報の提供を阻止するものである。

【００１５】

【発明が解決しようとする課題】

情報自動フィルタリングの大きな目的は不適切な情報を阻止する割合を増やすとともに、適切な情報が誤って阻止される割合を減らすことであるが、上述した各手法はそれぞれ一長一短があり、従来の情報自動フィルタリングでは十分なフィルタリング性能を得ることができないという問題がある。

【００１６】

具体的には、従来の自動フィルタリング手法では、例えば”Ｓｕｓｅｘ”というイギリスの町に関するホームページがブロックされるという悪例が報告されている。また、単語に対するスコアを利用する従来の方式では、単語および単語のスコアの設定がアドホックとなり、ユーザにとってどのように設定すれば最も有

効であるかに関して全く指針がなかった。そのため、提供を阻止すべき情報を阻止できなかったり、本来提供を阻止する必要のない情報が阻止されるなど、性能の点で問題があった。

【００１７】

例えば、「女子高生」という単語は一般的にポルノ情報に頻出すると考え、「女子高生」という単語とそのスコアを４０として登録したとする。その結果、「女子高生のサンプル画像、無料」という表現中に「女子高生」が含まれているため、この表現全体のスコアは４０となる。また、同様に「女子高生の乗ったバスが北海道で事故」という表現についてもこの表現全体のスコアは４０となり、これらの表現のスコアは同じになる。このため、閾値を２０としたとすると、本来阻止する必要のない後者の表現が阻止されてしまうという問題があり、また閾値を５０としたとすると、本来阻止すべき前者の表現が阻止されないという問題がある。これら２つの表現を区別するためには、「サンプル」「画像」「無料」などの単語や「バス」「北海道」「事故」といった単語にもスコアを設定する必要があることになるが、これらの単語は一般的にも良く利用される単語であり、スコアをどのように設定すれば良いかが明確でなく、スコアの設定により性能が大きく変動し、不適切な表現か否かの判定性能が十分に得られないという問題がある。

【００１８】

本発明は、上記に鑑みてなされたもので、その目的とするところは、単語の重みを容易かつ適確に設定し、この単語の重みを利用して情報が不適切であるか否かを適確に判定する情報自動フィルタリング方法および装置を提供することにある。

【００１９】

【課題を解決するための手段】

上記目的を達成するため、請求項１記載の本発明は、インターネットを介して提供される情報のうち不適切情報を識別し、この識別した不適切情報の提供を阻止する情報自動フィルタリング方法であって、提供の阻止を必要とする不適切な情報および提供の阻止を必要としない適切な情報を学習データとした自動学習に

より前記情報に含まれる単語に対して情報の提供を阻止する必要があるか否かを判定するために使用される単語の重みを求め、この求めた単語の重みを各単語に対応して重み付き単語リストとして記憶管理しておき、インターネットを介して提供される情報を入力し、この情報に含まれる単語を抽出し、この抽出した単語の各々に対する重みを前記重み付き単語リストから読み出し、この読み出した各単語の重みの総和を算出し、この算出した総和に基づき前記情報の提供を阻止すべきか否かを判定することを要旨とする。

【００２０】

請求項１記載の本発明にあっては、提供の阻止を必要とする不適切な情報および提供の阻止を必要としない適切な情報を学習データとした自動学習により単語の重みを求め、この求めた単語の重みを各単語に対応して重み付き単語リストとして記憶管理しておき、インターネットを介して提供される情報に含まれる単語を抽出し、この抽出した単語の各々に対する重みを重み付き単語リストから読み出し、この読み出した各単語の重みの総和を算出し、この総和に基づき前記情報の提供を阻止すべきか否かを判定するため、従来アドホックに設定しなければならなかった単語の重みを自動学習により適確に求め、この適確に求めた単語の重みを利用して情報が不適切な情報であるか否かを適確に判定し、不適切な情報の提供を阻止することができる。

【００２１】

また、請求項２記載の本発明は、請求項１記載の発明において、前記単語の重みを求める処理が、前記不適切な文書と適切な文書に対してベクトル空間上で弁別できる線形識別関数に基づく自動学習により単語の重みを求めることを要旨とする。

【００２２】

請求項２記載の本発明にあっては、不適切な文書と適切な文書に対してベクトル空間上で弁別できる線形識別関数に基づく自動学習により単語の重みを求めるため、単語の重みを適確に設定することができる。

【００２３】

更に、請求項３記載の本発明は、インターネットを介して提供される情報のう

ち不適切情報を識別し、この識別した不適切情報の提供を阻止する情報自動フィルタリング装置であって、提供の阻止を必要とする不適切な文書および提供の阻止を必要としない適切な文書を学習データとした自動学習により前記文書に含まれる単語に対して情報の提供を阻止する必要があるか否かを判定するために使用される単語の重みを求める単語重み学習手段と、この求めた単語の重みを各単語に対応して重み付き単語リストとして記憶管理する重み付き単語リスト格納手段と、インターネットを介して提供される情報を入力する入力手段と、この入力された情報に含まれる単語を抽出する単語抽出手段と、この抽出した単語の各々に対する重みを前記重み付き単語リストから読み出し、この読み出した各単語の重みの総和を算出し、この算出した総和に基づき前記情報の提供を阻止すべきか否かを判定する判定手段とを有することを要旨とする。

【００２４】

請求項３記載の本発明にあっては、提供の阻止を必要とする不適切な文書および提供の阻止を必要としない適切な文書を学習データとした自動学習により単語の重みを求め、この求めた単語の重みを各単語に対応して重み付き単語リストとして記憶管理しておき、インターネットを介して提供される情報に含まれる単語を抽出し、この抽出した単語の各々に対する重みを重み付き単語リストから読み出し、この読み出した各単語の重みの総和を算出し、この総和に基づき前記情報の提供を阻止すべきか否かを判定するため、従来アドホックに設定しなければならなかった単語の重みを自動学習により適確に求め、この適確に求めた単語の重みを利用して情報が不適切な情報であるか否かを適確に判定し、不適切な情報の提供を阻止することができる。

【００２５】

請求項４記載の本発明は、請求項３記載の発明において、前記単語重み学習手段が、前記不適切な文書と適切な文書に対してベクトル空間上で弁別できる線形識別関数に基づく自動学習により単語の重みを求める手段を有することを要旨とする。

【００２６】

請求項４記載の本発明にあっては、不適切な文書と適切な文書に対してベクト

ル空間上で弁別できる線形識別関数に基づく自動学習により単語の重みを求めるため、単語の重みを適確に設定することができる。

【００２７】

【発明の実施の形態】

次に、図１を参照して、本発明の実施形態に係る情報自動フィルタリング装置について説明する。同図に示す情報自動フィルタリング装置は、単語の重みを自動学習により求め、この自動学習で求めた単語の重みを利用して情報が不適切であるか否かを判定し、不適切な情報の提供を阻止するものであり、インターネットを介して提供されるＨＴＭＬ情報を入力する入力部１、この入力部１を介して入力された情報に出現する単語抽出部３、提供の阻止を必要とする不適切な情報である文書および提供の阻止を必要としない適切な情報である文書を学習データとした自動学習により前記文書に含まれる単語に対して情報の提供を阻止する必要があるか否かを判定するために使用される単語の重みを求める重み付き単語リスト学習部６０、この重み付き単語リスト学習部６０で求めた単語の重みを各単語に対応して重み付き単語リストとして記憶管理する重み付き単語リスト格納部５０、単語抽出部３で抽出された単語および該単語に対して重み付き単語リスト格納部５０から得られた単語の重みｗに基づき入力部１から入力された情報の提供を阻止すべきか否かを判定する自動フィルタリング部３０、および該自動フィルタリング部３０で得られた判定結果を出力する出力部４０から構成されている。

【００２８】

本実施形態の情報自動フィルタリング装置は、重み付き単語リスト学習部６０において単語の重みを自動学習により予め取得し、この自動学習で得た単語の重みを利用することを特徴とする。この単語の重みの自動学習の方法を図２のフローチャートに示す単語重みの学習アルゴリズムで行われるものである。すなわち、図２に示す学習アルゴリズムでは、学習データの集合Ｅ＝｛ｄ１，…，ｄｎ｝として提供の阻止を必要とする不適切な情報および提供の阻止を必要としない適切な情報を重み付き単語リスト学習部６０に入力し、この入力された不適切な情報と適切な情報をベクトル空間上で弁別する線形識別関数から単語の重みを取得

する。具体的には次ぎの手順で行う。

【００２９】

まず、入力部１から入力されたＨＴＭＬ文書をベクトル空間モデルによって表現する。すなわち、すべての文書を表現するｎ個の単語を選択し、それぞれの文書をｎ次元のベクトルで次式のように表現する。

【００３０】

【数１】

$$\vec{V}_d = (f_{d1}, \cdots, f_{di}, \cdots, f_{dn}) \qquad \cdots (1)$$

このベクトルの各要素は、各々単語の文書ｄでの出現頻度を正規化したものである。単語の出現頻度の正規化には次に示す数式で表されるＴＦ＊ＩＤＦという手法を用いている。

【００３１】

【数２】

$$f_{di} = tf_{di} * \log\left(\frac{N}{df_i}\right) \qquad \cdots (2)$$

ここで、ｔｆｄｉは単語ｉが文書ｄに出現する頻度、Ｎはすべての文書の数、ｄｆｉは単語ｉが出現する文書の数である。

【００３２】

自動フィルタリングは、次に示す数式で表される線形識別関数によって行われ、この関数によって単語の重みの総和Ｄｉｓ（ｄ）が計算される。

【００３３】

【数３】

$$Dis(d) = \sum_{i=1}^{n} w_i * f_{di} \qquad \cdots (3)$$

ここで、ｗｉは各単語ｉに対する重みであり、ｆｄｉは上式（２）の値であり、文書における各単語のｆｄｉ値である。

【００３４】

上述した式（３）から、総和Ｄｉｓ（ｄ）が０より大きい場合、前記文書は有害であり、０以下である場合、無害であると判定される。

【００３５】

なお、上述した各単語ｉに対する重みは文書ｄが有害な場合、総和Ｄｉｓ（ｄ）＞０となり、無害な場合、総和Ｄｉｓ（ｄ）≦０となるように設定される。

【００３６】

次に、この単語の重みの学習アルゴリズムについて図２に示すフローチャートを参照して説明する。なお、この単語の重みの学習には perceptron learning algorithm（ＰＬＡ）を使用している。

【００３７】

図２においては、まず各種パラメータを設定する（ステップＳ５１）。このパラメータとしては、各単語の重みの集合Ｗ＝（ｗ１，…，ｗｎ）、Ｎ個の学習データＥ＝｛ｄ１，…，ｄｎ｝、定数η、最大学習回数Ｍａｘ、図２に示す学習処理を繰り返し行う学習回数ｍがある。

【００３８】

それから、全ての文書を表現する単語のうち頻度の高いｎ個の単語を選択する（ステップＳ５２）。

【００３９】

次に、単語の重みの集合Ｗを初期化する（ステップＳ５３）。この初期化では、各単語の重みに乱数を入力する。それから、すべての学習データに対して前記単語重みの総和Ｄｉｓ（ｄ）を上式（３）により計算する（ステップＳ５５）。

【００４０】

そして、この計算の結果、すべての無害な文書ｄについて総和Ｄｉｓ（ｄ）≦０であり、かつすべての有害な文書ｄについて総和Ｄｉｓ（ｄ）＞０であるか否かをチェックし（ステップＳ５７）、そうである場合には、処理を終了するが、そうでない場合には、このように誤って分類されたすべての文書ｄについて次のステップＳ６１，Ｓ６３で示すように重みの変化度合Ｓを補正する（ステップＳ５９）。

【００４１】

すなわち、ステップＳ６１では、文書ｄｉが有害であって、かつ総和Ｄｉｓ（ｄ）≦０の場合には、重み変化度合Ｓを増加するように補正し、またステップＳ６３では、文書ｄｉが無害であって、かつ総和Ｄｉｓ（ｄ）＞０の場合には、重み変化度合Ｓを低減するように補正する。

【００４２】

そして、このように補正された重み変化度合Ｓを使用して単語重みの集合ＷをステップＳ６５で示す式のように補正する。それから、学習回数ｍを＋１インクリメントし（ステップＳ６７）、この学習回数ｍが最大学習回数Ｍａｘより小さいか否かをチェックし（ステップＳ６９）、また最大学習回数Ｍａｘより小さい場合には、ステップＳ５５に戻り、ステップＳ５７に示した条件が満たされるまで、ステップＳ５５以降の処理を繰り返し行う。そして、最終的にｎ個の単語に対する単語重みの集合が求まる。

【００４３】

重み付き単語リスト学習部６０で取得された各単語の重みは、各単語に対応して重み付き単語リストとして重み付き単語リスト格納部５０に格納される。次に示す表７は、重み付き単語リスト格納部５０に格納されている重み付き単語リストを示す表であり、各単語に対応して単語重みｗが格納されている。

【００４４】

【表１】

重み付き単語リスト	
単　語	単語重みｗ
⋮	
画　像	１０．９
⋮	⋮
サンプル	１８．７
⋮	⋮
事　故	－１６．６
⋮	⋮
女子高生	８２．２
⋮	⋮
バ　ス	－１０１．９
⋮	⋮
北　海　道	－１１２．５
⋮	⋮
無　料	－６．３
⋮	⋮

次に、このように重み付き単語リスト学習部６０で得られ、重み付き単語リスト格納部５０に格納された単語重みに基づきインターネットから提供された情報が不適切な情報であるか否かを判定する処理について説明する。

【００４５】

図１において、入力部１から入力されたインターネットからの情報は、単語抽出部３で、重み付き単語リスト格納部５０に格納されている単語リストと照合し、入力情報中に出現する単語とその出現頻度を求める。また、同時に出現した単語の重みｗも重み付き単語リスト格納部５０から求め、出現単語とその頻度および重みを自動フィルタリング部３０に供給する。自動フィルタリング部３０は、この入力された単語に対する重みｗと出現頻度から、入力情報中に出現した全て

の単語に対する重みｗの総和を算出し、この総和を所定の閾値と比較し、総和が閾値よりも大きい場合不適切な情報と判定し、総和が閾値よりも小さい場合、適切な情報と判定し、この判定結果を出力部４０から出力する。

【００４６】

具体的に説明する。表１に示すように、重み付き単語リスト学習部６０では、予め入力された学習データから「画像」の重みは１０．９、「サンプル」の重みは１８．７、「事故」の重みは－１６．６、「女子高生」の重みは８２．２、「バス」の重みは－１０１．９、「北海道」の重みは－１１２．５、「無料」の重みは－６．３と求まり、重み付き単語リスト格納部５０に格納しているので、この結果を利用すると、例えば「女子高生の乗ったバスが北海道で事故」という表現全体に対しては、自動フィルタリング部３０で各単語の重みの総和を求め、８２．２－１０１．９－１１２．５－１６．６＝－１４８．８となる。また、「女子高生のサンプル画像、無料」の表現全体に対しては、自動フィルタリング部３０で各単語の総和を求め、８２．２＋１８．７＋１０．９－６．３＝１０５．５となる。そして、図２の処理と同様に閾値を０とすると、「女子高生の乗ったバスが北海道で事故」という表現は閾値を下回るので、情報の提供は阻止されず、また「女子高生のサンプル画像、無料」という表現は閾値を上回るので、情報の提供は阻止されるというように正しく判定することができる。

【００４７】

次に、図４および図６を参照して、本発明の他の実施形態に係る自動フィルタリング装置について説明する。図４に示す自動フィルタリング装置は、図６で説明した学習により単語リストを作成する情報自動フィルタリング装置２５に対して第三者判定フィルタリング処理部２３および該第三者判定フィルタリング処理部２３で有害ＵＲＬを参照するために使用される有害ＵＲＬ一覧テーブル格納部１７が付加されている。

【００４８】

有害ＵＲＬ一覧テーブル格納部１７は、有害情報を提供するＵＲＬを有害ＵＲＬ一覧テーブルとして格納しているものであり、第三者判定フィルタリング処理部２３は、前記入力部１から入力されたＨＴＭＬ文書のＵＲＬを有害ＵＲＬ一覧

テーブル格納部１７の有害ＵＲＬ一覧テーブルに登録されている各ＵＲＬと照合し、一致するＵＲＬがあるか否かを判定するものである。

【００４９】

図６は、図４に示す自動フィルタリング装置の更に詳細な構成を示すブロック図である。図６に示す自動フィルタリング装置は、図６に示した学習により作成した重み付き単語リストを用いた情報自動フィルタリング装置を構成する入力部１、単語抽出部３、重み付き単語リスト格納部５０、自動フィルタリング部３０、出力部４０に加えて、図４の第三者判定フィルタリング処理部２３に対応するＵＲＬリストに基づくフィルタリング部１５および有害ＵＲＬ一覧テーブル格納部１７を有している。

【００５０】

このように構成される自動フィルタリング装置、すなわち第三者判定フィルタリング処理部によるＵＲＬリスト一覧と学習により作成した重み付き単語リストを用いた情報自動フィルタリング装置によるフィルタリング処理では、まずインターネット２１を介して入力されたＨＴＭＬ文書は、そのＵＲＬが有害ＵＲＬ一覧テーブル格納部１７の有害ＵＲＬ一覧テーブルに登録されている各ＵＲＬと照合され、一致するＵＲＬがあるか否かが判定される。そして、有害ＵＲＬ一覧テーブル格納部１７の有害ＵＲＬ一覧テーブルに登録されたＵＲＬと一致する場合には、このＵＲＬが示す情報の提示は阻止される。

【００５１】

ＵＲＬリストに基づくフィルタリング部１５による有害ＵＲＬ一覧テーブルを参照した判定の結果、有害ＵＲＬ一覧テーブル格納部１７の有害ＵＲＬ一覧テーブルに登録されているＵＲＬと一致するものがない場合には、学習により作成した重み付き単語リストを用いた情報自動フィルタリング装置２５によるフィルタリングが図６で説明したように行われる。

【００５２】

このように本実施形態では、第三者による判定に基づくフィルタリングと学習により作成した重み付き単語リストを用いたフィルタリングの両方が行われるため、有害情報を適確に検出して阻止することができる。

【００５３】

【発明の効果】

以上説明したように、本発明によれば、提供の阻止を必要とする不適切な情報および提供の阻止を必要としない適切な情報を学習データとした自動学習により単語の重みを求め、この単語の重みを各単語に対応して重み付き単語リストとして記憶管理し、インターネットを介して提供される情報に含まれる単語を抽出し、この抽出した単語の各々に対する重みを重み付き単語リストから読み出し、各単語の重みの総和を算出し、この総和に基づき情報の提供を阻止すべきか否かを判定するので、従来アドホックに設定しなければならなかった単語の重みを自動学習により適確に求め、この適確に求めた単語の重みを利用して情報が不適切な情報であるか否かを適確に高い性能で判定し、不適切な情報の提供を阻止することができる。

【図面の簡単な説明】

【図１】

本発明の別の実施形態に係る情報自動フィルタリング装置の構成を示すブロック図である。

【図２】

図１に示すフローチャートに使用されている単語重みの設定手順を示すフローチャートである。

【図３】

本発明の他の実施形態に係る自動フィルタリング装置の概要構成を示す説明図である。

【図４】

従来の自己判定に基づくフィルタリングを説明するための図である。

【図５】

図４に示した自己判定に基づくフィルタリングの一例としてRSACi とSafeSurfによる評価結果の記述例を示す図である。

【図６】

従来の第三者による判定に基づく有害情報フィルタリングを説明するための図

である。

【符号の説明】

１　入力部

３　単語抽出部

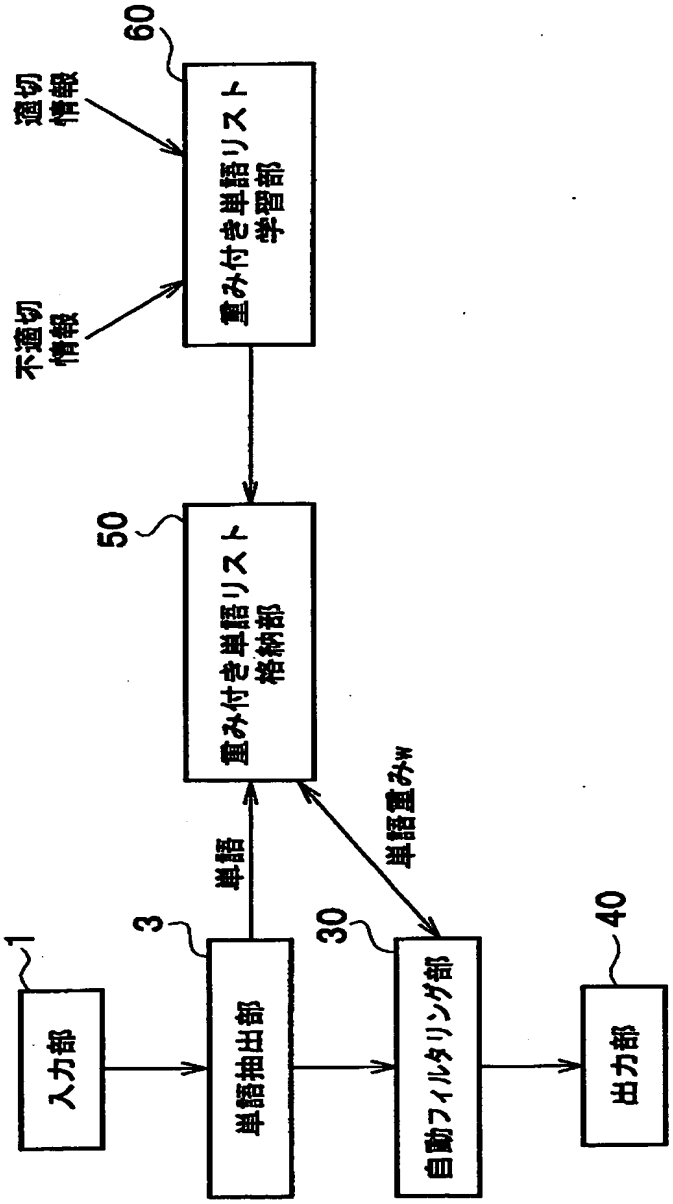
３０　自動フィルタリング部

５０　重み付き単語リスト格納部

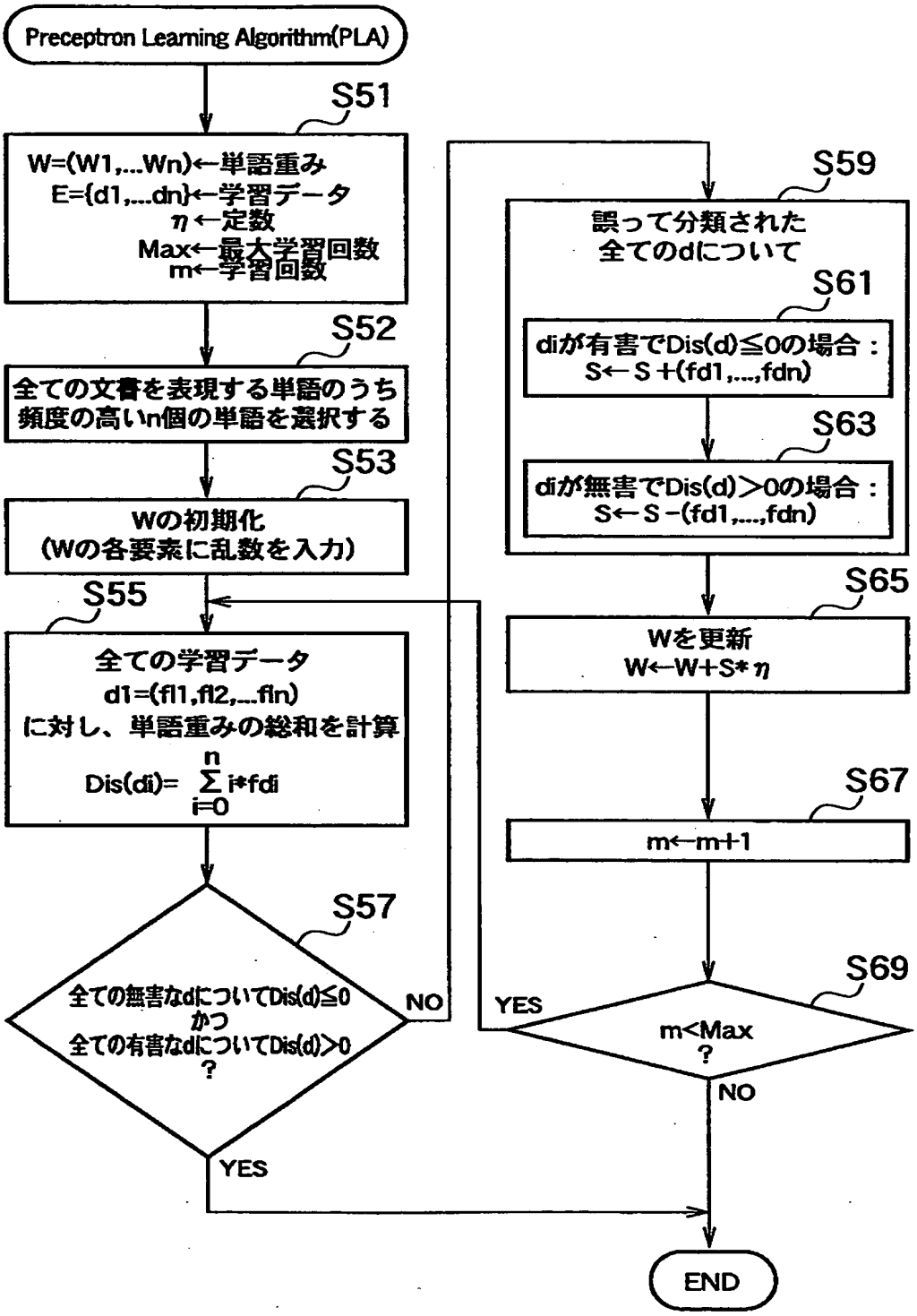
６０　重み付き単語リスト学習部

【書類名】　　図面

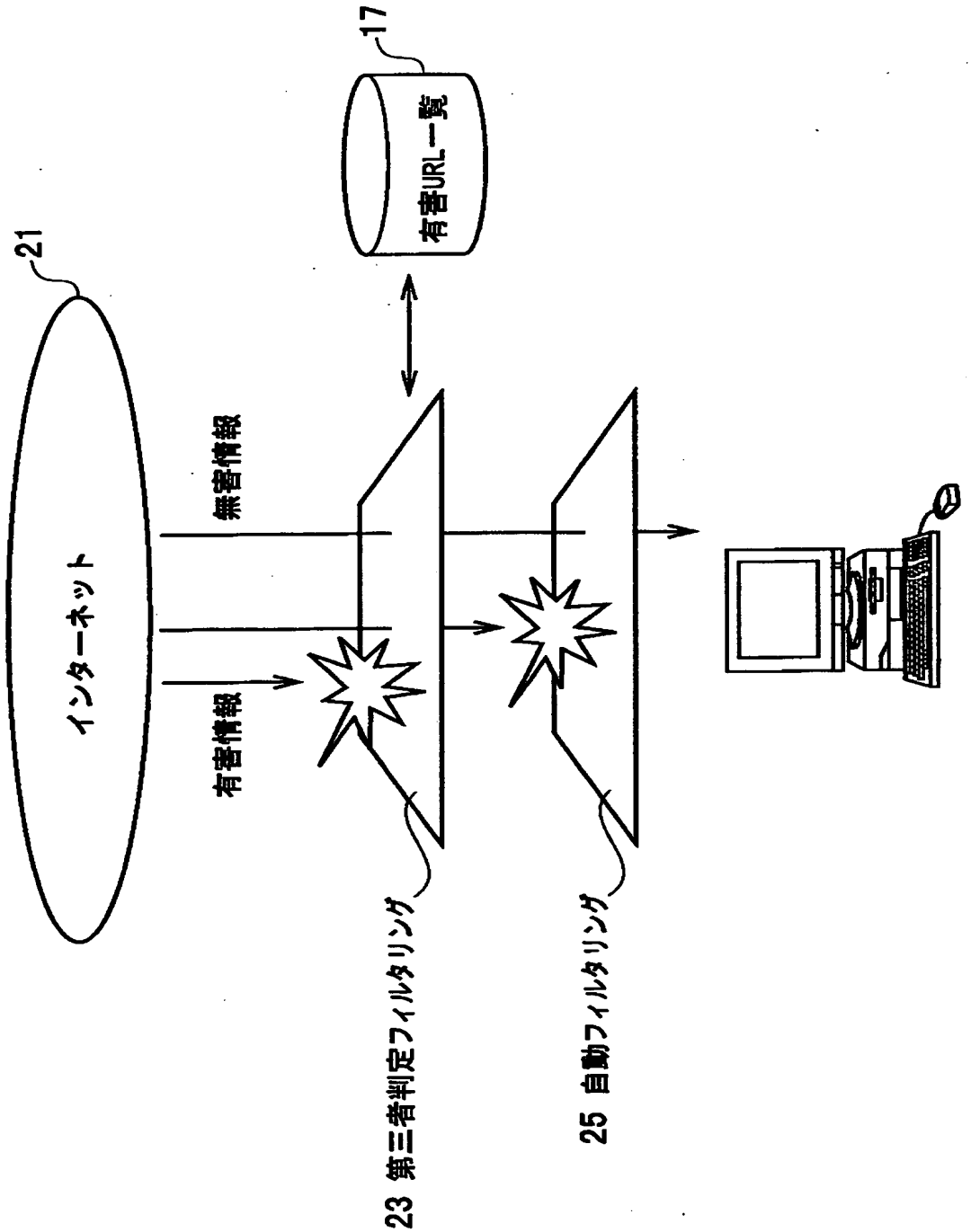
【図１】


60
適切情報
不適切情報
重み付き単語リスト学習部
50
重み付き単語リスト格納部
単語
単語重みw
3
30
40
1
入力部
単語抽出部
自動フィルタリング部
出力部

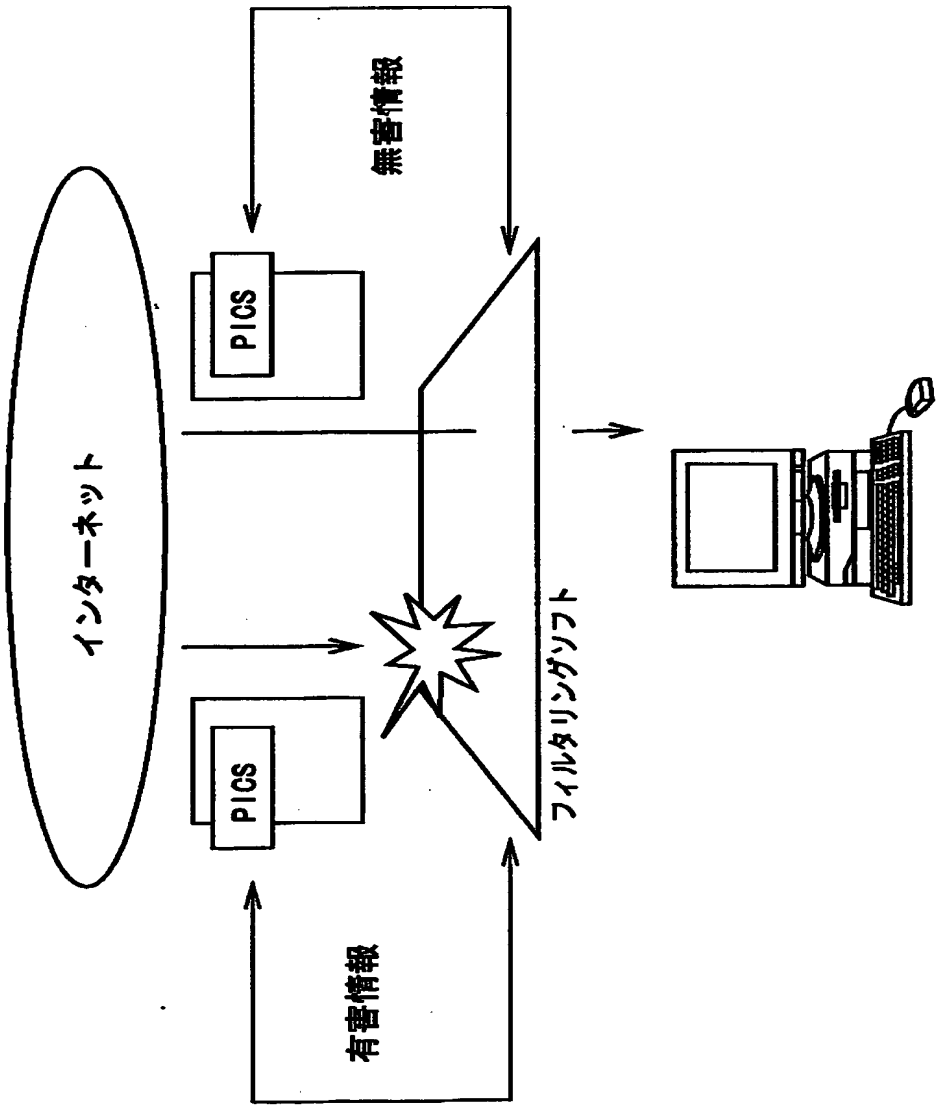
【図２】


Preceptron Learning Algorithm(PLA)
S51
W=(W1,...Wn)←単語重み
E={d1,...dn}←学習データ
η←定数
Max←最大学習回数
m←学習回数
S52
全ての文書を表現する単語のうち
頻度の高いn個の単語を選択する
S53
Wの初期化
(Wの各要素に乱数を入力)
S55
全ての学習データ
d1=(fl1,fl2,...fln)
に対し、単語重みの総和を計算
Dis(di)= Σ i=0 n i*fdi
S57
全ての無害なdについてDis(d)≦0
かつ
全ての有害なdについてDis(d)>0
?
NO
YES
S59
誤って分類された
全てのdについて
S61
diが有害でDis(d)≦0の場合：
S←S+(fd1,...,fdn)
S63
diが無害でDis(d)>0の場合：
S←S-(fd1,...,fdn)
S65
Wを更新
W←W+S*η
S67
m←m+1
S69
m<Max
?
YES
NO
END

【図３】


17
有害URL一覧
21
インターネット
無害情報
有害情報
23 第三者判定フィルタリング
25 自動フィルタリング

【図４】

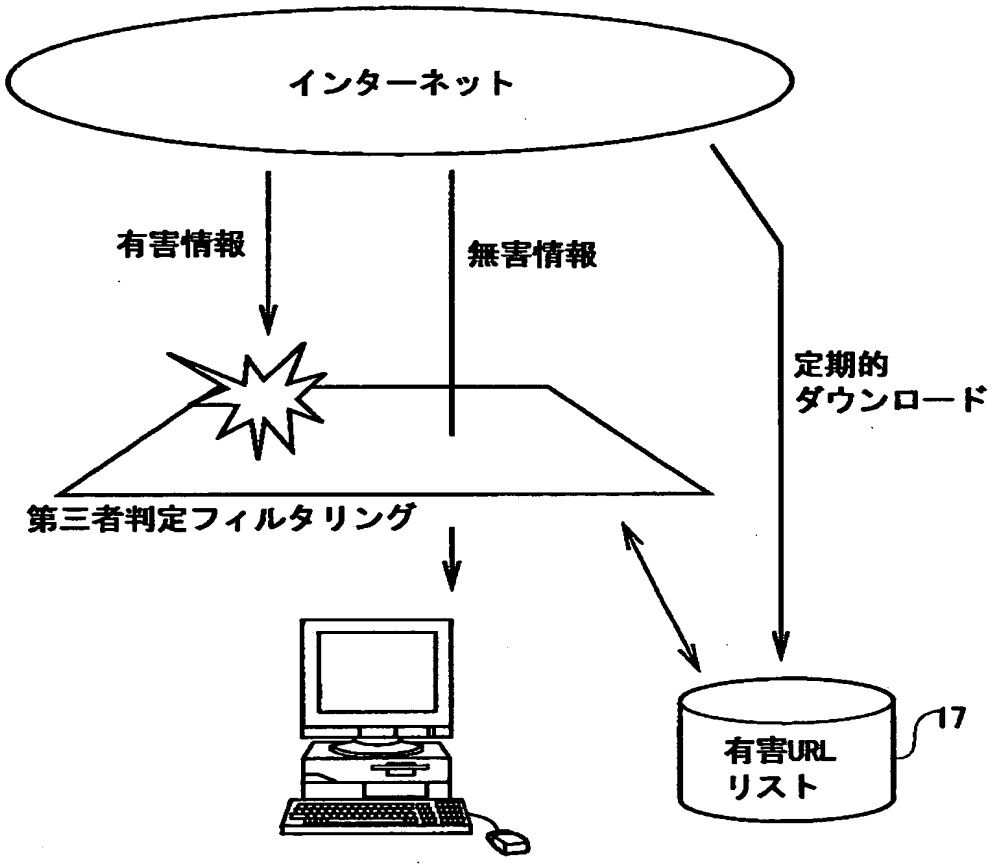


【図５】

```
〈HTML〉
〈HEAD〉
〈TITLE〉
Example of RSAC1 and SafeSurf Rated Page
〈/TITLE〉

〈META http-equiv-"PICS-Label" content- '(PICS-1.0
"http://www.rsac.org/ratingav01.html"
1 gen true comment "RASC1 North Amerika Serber" by
"marks@icrosys.com" for "http://www.samplesite.com" on
"1996.09.05T02:15-0800" exp "1997.09.05T02:15-0800"
r (n 2 s 0 v 1 1 2))'〉

〈META http-equiv-"PICS-Label" content- '(PICS-1.1
"http://www.classify.org/safesurf/" 1 gen t for
"http://www.samplesite.com/"
r(SS--000 2 SS--100 1))'〉

〈/HEAD〉
〈BODY〉
...
```

RSACiとSafeSurfによる評価結果の記述例

【図６】



【書類名】　　要約書

【要約】

【課題】　単語の重みを容易かつ適確に設定し、この単語の重みを利用して情報が不適切であるか否かを適確に判定する情報自動フィルタリング方法および装置を提供する。

【解決手段】　重み付き単語リスト学習部６０に学習データとして提供の阻止を必要とする不適切な情報と提供の阻止を必要としない適切な情報を入力し、不適切な情報と適切な情報をベクトル空間上で弁別する線形識別関数から単語の重みを取得して重み付き単語リストとして重み付き単語リスト格納部５０に格納し、入力部１からの情報から単語抽出部３で単語を抽出し、この単語の重みｗを重み付き単語リスト格納部５０から取得して自動フィルタリング部３０に入力し、これらの単語の重みｗの総和を算出し、総和が閾値よりも大きい場合不適切な情報と判定し、総和が閾値よりも小さい場合、適切な情報と判定する。

【選択図】　　図１

出　願　人　履　歴　情　報

識別番号　　〔０００００１２１４〕

１．変更年月日　　１９９８年１２月　３日
［変更理由］　名称変更
住　所　東京都新宿区西新宿２丁目３番２号
氏　名　ケイディディ株式会社